MockData

Adriel Evaristo

2025-05-05

Mock Dataset

```
# Load libraries
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.2     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
set.seed(42)

# Dataset: 4 cities, 20 mice each
mock_data <- data.frame(
  mouse_id = 1:80,
  city = rep(c("San Francisco", "Oakland", "San Jose", "San Rafael"), each = 20),
  body_weight = c(rnorm(20, mean = 22, sd = 2),   # San Francisco
                  rnorm(20, mean = 21, sd = 2),   # Oakland
                  rnorm(20, mean = 23, sd = 2.5), # San Jose
                  rnorm(20, mean = 20, sd = 1.8)) # San Rafael
)

head(mock_data)
```

```
##   mouse_id          city body_weight
## 1        1 San Francisco    24.74192
## 2        2 San Francisco    20.87060
## 3        3 San Francisco    22.72626
## 4        4 San Francisco    23.26573
## 5        5 San Francisco    22.80854
## 6        6 San Francisco    21.78775
```

```r
write.csv(mock_data, "mock_body_weight_data.csv", row.names = FALSE)
```

Create Plot

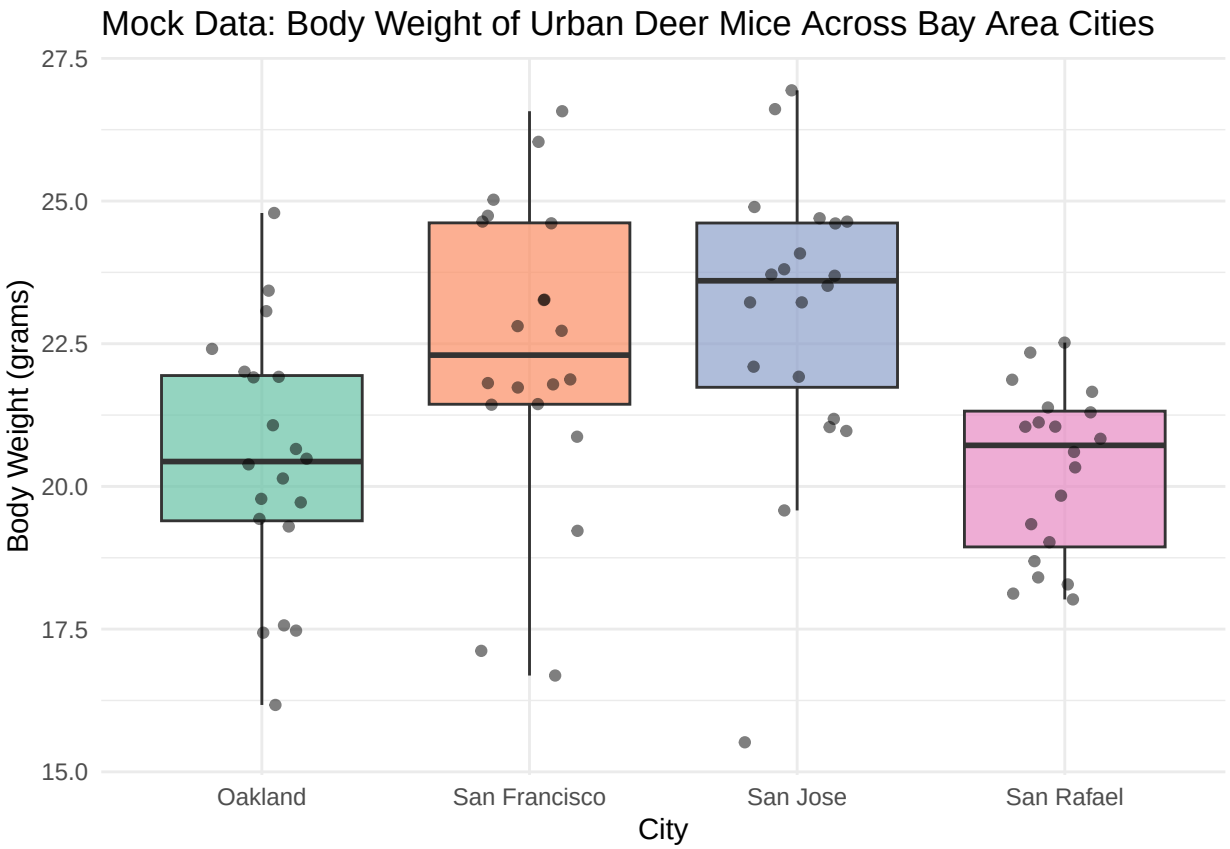
```r
# Read the mock data
data <- read.csv("mock_body_weight_data.csv")

# One-way ANOVA
anova_result <- aov(body_weight ~ city, data = data)
summary(anova_result)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## city          3  111.4   37.15   7.246 0.000244 ***
## Residuals    76  389.6    5.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Tukey
TukeyHSD(anova_result)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = body_weight ~ city, data = data)
##
## $city
##                                diff         lwr        upr     p adj
## San Francisco-Oakland     1.9258236  0.04503137  3.8066159 0.0427195
## San Jose-Oakland          2.5397068  0.65891452  4.4204991 0.0036775
## San Rafael-Oakland       -0.1689351 -2.04972739  1.7118571 0.9953380
## San Jose-San Francisco    0.6138832 -1.26690911  2.4946754 0.8266414
## San Rafael-San Francisco -2.0947588 -3.97555102 -0.2139665 0.0229892
## San Rafael-San Jose      -2.7086419 -4.58943418 -0.8278496 0.0017140
```

```r
#Boxplot
ggplot(data, aes(x = city, y = body_weight, fill = city)) +
  geom_boxplot(outlier.shape = NA, alpha = 0.7) +
  geom_jitter(width = 0.2, alpha = 0.5) +
  scale_fill_brewer(palette = "Set2") +
  labs(
    title = "Mock Data: Body Weight of Urban Deer Mice Across Bay Area Cities",
    x = "City",
    y = "Body Weight (grams)"
  ) +
  theme_minimal() +
  theme(legend.position = "none")
```


Mock Data: Body Weight of Urban Deer Mice Across Bay Area Cities
Body Weight (grams)
27.5
25.0
22.5
20.0
17.5
15.0
Oakland
San Francisco
San Jose
San Rafael
City